



Intézményi repozitóriumok és digitális megőrzés

Előzmények

2010 tavaszán a *University of Massachusetts Amherst* két könyvtári munkatársa egy országos felmérést végzett: szerettek volna képet kapni arról, hogy az *ARL (Association of Research Libraries)* tagkönyvtáraiban milyen módon oldják meg az intézményi repozitóriumban tárolt digitális tartalmak hosszú távú megőrzését.

A digitális megőrzés komoly probléma, amelyet a könyvtáraknak kezelniük kell. Nagy kérdés, hogy hogyan lehet a távoli jövő számára is átmenteni azt a digitális információt, amelyet egyszerre fenyeget a technológiai elavulás és a tárolóeszköz tönkremenetele. Míg egy kézírásos kódex akár ezer évig is fennmarad, addig egy átlagos CD 15 év alatt tönkremegy, sőt az is lehet, hogy a rajta tárolt állományok formátuma még hamarabb elavul. *Paul Conway* 1996-os cikkében [1] így jellemezte ezt az ellentmondásos helyzetet: miközben információtároló eszközeink kapacitása exponenciálisan növekszik, ezek élettartama ugyanilyen ütemben csökken. *Julian Jackson* „*Digital Longevity*” c. írásában [2] pedig arra hívta fel a figyelmet, hogy a számítástechnika gyors változásai miatt akár egy évtizeden belül is megtörténhet, hogy nem lehet többé megjeleníteni egyes információs állományokat. Ahhoz, hogy a digitálisan tárolt tudományos eredmények megmaradjanak a jövőnek, a könyvtáraknak – a szerzőkkel együttműködve – lépéseket kell tenniük a megfelelő megőrzési módszerek kialakítása érdekében. Az egyik ilyen, viszonylag új fejlemény az intézményi repozitórium (IR) megjelenése és elterjedése, ahová a szerzők archiválás céljából beadhatják az anyagaikat. Kérdés persze, hogy az IR alkalmas-e a hosszú távú megőrzésre? Egy 2007-es kutatás jelentésének előszavában [3] *Abby Smith* így foglalta össze az ezzel kapcsolatos tapasztalatait: nincs konszenzus arról, hogy mire is szolgál az IR; sok intézmény elsősorban a tudományos kommunikáció elősegítésére hozta létre a repozitóriumát, és csupán egy

részük kötelezte el magát a beadott tartalom gondozása és megőrzése mellett. Ugyanakkor *Clifford Lynch*, akinek az IR fogalmára adott definícióját gyakran idézik [4], a repozitóriumok egyik alapvető feladatáknak határozta meg a hosszú távú megőrzést és felügyeletet, a beadott tartalom rendszerezése és szolgáltatása mellett.

Nemcsak maguk az IR-üzemeltetők, hanem az ezzel a „műfajjal” foglalkozó kutatók is aránylag kevés figyelmet fordítottak korábban a megőrzés problematikájára. Ifj. *Charles W. Bailey* az *Institutional Repository Bibliography* [5] összeállítója alig néhány publikációt tudott csak felvenni a *Digital Preservation* rovatba. Ugyanakkor a hosszú távú megőrzés szükségességét még jóval az IR feltalálása előtt megfogalmazták többen is. *Don Waters* és *John Garrett* például már 1996-ban ezt írták mérföldkőnek számító jelentésükben [6]: „Ha nem találunk megbízható eszközöket és módszereket a digitális megőrzésre, ennek garantáltan súlyos és hosszú időre szóló kulturális következményei lesznek.” Nagy-Britanniában még ebben az évben létrejött a *Digital Preservation Coalition*, az USA-ban pedig a *Library of Congress* egy országos stratégiát dolgozott ki. A *CCSDS (Consultative Committee for Space Data Systems)* 2002-ben közreadta az *OAIS* modell leírását (*Recommendation for Space Data System Standards Reference Model for an Open Archive Information System*), mely egy, a digitális megőrzés valamennyi fázisára kiterjedő elméleti keretet határoz meg: bevételezés, tárolás, visszakeresés és hosszú távú archiválás. Ugyanakkor az eddigi felmérések azt jelzik, hogy a már működő IR-ek esetében ez utóbbi funkció ellátása eléggé gyerekcipőben jár. A már korábban említett 2007-es jelentés [3] szerint még nagy a távolság a szándék és a tényleges gyakorlat/képesség között. 2005-ben egy másik kutatás [7] azzal az eredménnyel zárult, hogy csak az intézmények egyharmadában van megtervezett, jóváhagyott és megvalósított digitális megőrzési politika.

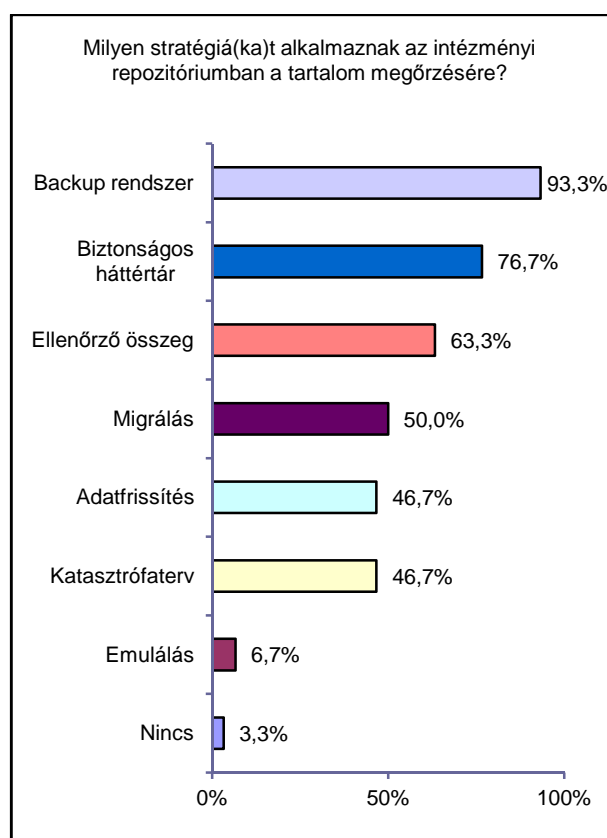
Bár az IR rendszerekbe egyre több ilyen jellegű funkciót is beépítenek, a digitális megőrzés nem oldható meg csupán szoftveres úton – természetes és jogi személyek, intézmények, számítógépes eszközök és alkalmazások, valamint munkafolyamatok együttműködésére, illetve összehangolására van hozzá szükség. Ha nincs elegendő kontroll a beadott tartalom felett, ha nem optimális formátumban tárolják az állományokat, ha kevés a metaadat, vagy ha nincsenek tisztázva a szellemi tulajdonjogok, akkor ezek a tényezők megnehezítik, illetve megdrágítják a hosszú távú archiválást.

A felmérés ismertetése

A cikk szerzői az ARL 125 tagja közül annak a 72 tudományos könyvtárnak küldték ki a kérdőívüket, amelyek saját IR-t működtetnek – abban a reményben, hogy ezeken a helyeken legalább elméletben foglalkoztak már a hosszú távú megőrzés kérdésével. A megkérdezettek 52 százaléka válaszolt, közülük 43 százalék töltötte ki teljesen a kérdőívet. A kutatók összesen 24 kérdést tettek fel, melyek a megőrzési stratégiára és gyakorlatra, a jogi helyzetre, a megőrzendő tartalom minőségére és a fenntarthatóságra vonatkoztak.

A válaszadó könyvtárak 57,9 %-a a DSpace-t használja, a többinél a Digital Commons, a ContentDM, a DigiTool, illetve valamilyen egyéb rendszer képezi a repozitórium műszaki hátterét. 97,4 százalékuk mondta azt, hogy a megőrzés volt az egyik cél, ami miatt az IR-t létrehozták, és az intézményeknek valamivel több mint a fele írott szabállyal is rendelkezik erre vonatkozóan. Nagy többségük csak bizonyos fájlformátumok megőrzését vállalja fel hosszú távon (pl. PDF, RTF, ODT, XML, TIFF, JPEG2000, AIFF, AVI). Arra a kérdésre, hogy milyen módszereket használnak, a válaszok megoszlása a mellékelt grafikonon látható (1. ábra). A legtöbb helyen megelégszenek a megbízható háttértárra történő időnkénti biztonsági mentésekkel, illetve az IR rendszerbe beépített *checksum* funkcióval, amely a fájlok integritásának ellenőrzésére szolgál. A munkaidényesebb és költségesebb megoldásokat – mint amilyen az egyik rendszerből vagy formátumról egy másikra való migrálás, vagy a más adathordozóra való átmásolással történő frissítés, illetve a korábbi szoftverkörnyezet emulálása – csak a válaszadók fele vagy még kisebb hányada jelölte meg. Amellett, hogy kihasználják az IR rendszerük egyes beépített megőrzési funkcióit, a könyvtárak 65,5 százaléka külső szolgáltatásokat vagy eszközöket is igénybe vesz. Ilyenek

például a LOCKSS, a MetaArchive, a DuraCloud, az iRODS, illetve a CDL kurátori szolgáltatásai, az InterPARES és a Bepress biztonsági mentései, és persze az egyetemük saját háttértárai.



1. ábra A tartalom őrzésének stratégiái

Biztató jel, hogy 58,6 százalék válaszolt igennel arra a kérdésre, hogy tárolnak-e az IR rendszerükben megőrzési típusú metaadatokat is. Ezek leggyakrabban a technikai jellemzőkre, a jogi státuszra, valamint a származási helyre, illetve az eredeti tulajdonosra vonatkozó információk, továbbá a digitális objektumon történt esetleges változtatások leírásai. Szinte valamennyi rendszerrel van lehetőség a teljes tartalom és a metaadatok exportálására, ami egy jövőbeli új rendszerbe való áttölthetőség szempontjából fontos feltétel.

Az intézményi repozitóriumok általában az OA (*open access*) elvén működnek, vagyis az archívumba feltöltők engedélyezik az IR üzemeltetői számára, hogy nyilvános hozzáférést biztosítsanak az általuk beadott tartalomhoz. Viszont további feladatot jelent a hosszú távon való megőrzéshez és szolgáltatáshoz szükséges jogok és engedélyek megszerzése, mert ennek megvalósításához elkerülhetetlen lehet például a fájlok más formátumba

való konvertálása, vagy akár a digitális tartalom bizonyos mértékű módosítása is. Ehhez nem biztos, hogy elegendő a beadók hozzájárulása, hanem szükség lehet az eredeti copyright-tulajdonosok (pl. kiadók) engedélyére is. A válaszolók 72,4 százaléka jelezte azt, hogy mindig kérnek hozzájárulást a beadóktól a megőrzésre vonatkozóan is (pl. a beadási űrlapon kiválasztható opció, írott szerződés, szabályzat, egyetértési nyilatkozat, vagy szóbeli beleegyezés formájában), de csak 56,7 százalékuk szokott ilyet beszerezni az eredeti jogtulajdonostól is, ha az nem azonos a beadóval.

A megőrzés szempontjából fontos kérdés a megőrzendő tartalom minősége is, amely pedig nagyban függ annak gyűjtési módjától. Az ellenőrizetlen tartalomfeltöltés könnyen vezethet gyorsan elavuló formátumokhoz, gyenge minőséghez, olvashatatlan szkennelésekhez, vagy hiányos metaadatokhoz, melyek azután megnehezítik vagy értelmetlenné teszik a hosszú távú megőrzésre irányuló erőfeszítéseket. A vizsgált IR-ek nyolcvan százalékában van valamiféle gyarapítási szabályzat. Ezek leginkább a beadható anyagok témájára és típusára, illetve a fájlformátumokra vonatkozóan fogalmaznak meg kritériumokat, továbbá egyes funkciókat (pl. visszavonás, hozzáférés, megőrzés) szabályoznak. A legtöbb helyen mindhárom lehetséges beadási mód előfordul: önarchiválás a szerzők részéről, közvetítőn keresztül történő beszállítás, valamint a repozitóriumot fenntartók által végzett tartalombővítés. Arányuk nagyon változó, de általában a harmadik a leggyakoribb módszer. Ami a minőségbiztosítást illeti: 83,3 százalék válaszolta azt, hogy van autentikáció a rendszerükben, vagyis az adminisztrátorok azonosítani tudják, hogy ki és mit töltött fel az IR-be. 70 százalék azok aránya, akik beadási útmutatóval próbálják segíteni és irányítani a feltöltőket, a válaszadó könyvtárak kétharmadánál pedig az üzemeltetők át is nézik a beadott anyagokat, viszont csak egyötödük jelölte be a lektorálást mint a minőségbiztosítás egyik formáját. De feltehetően ezek többsége sem saját *peer review* rendszert működtet, hanem egyszerűen arról van szó, hogy folyóiratokban megjelent, így már egy lektorálási folyamaton átment cikkeket vesznek fel a repozitóriumba.

A kérdőív utolsó része az IR fenntarthatóságára vonatkozott, mert ez közvetlenül érinti a benne tárolt tartalom megőrzését. Bár a visszaérkezett válaszok 63,3 százaléka szerint náluk megoldottnak látszik hosszabb távon is a rendszer fenntartása, elég nagy azon könyvtárak aránya, ahol nincs elég forrás az IR üzemeltetésére, vagy bi-

zonytalan annak jövője. A pénzügyi háttér mellett a megfelelő képzettségű és számú személyzet is kulcskérdés. Ezen a téren nagyon megoszlottak az eredmények: csak a könyvtárak felénél van elegendő szakember erre a feladatra, másik felénél kevésnek ítélik a munkaerőt (van, ahol csak egy főállású dolgozó lett ezzel a munkával megbízva), különösen annak fényében, hogy a beérkező és megőrzendő tartalom egyre jobban növekszik.

A záró kérdések a megőrzés tervezett időtartamát firtatták. Az IR üzemeltetők 20 százaléka rövid távú megőrzésre rendezkedett be, tehát csak addig akarják szolgáltatni az egyes digitális dokumentumokat, amíg azokra várhatóan igény van, illetve amíg nem avulnak el egy technológiaváltás miatt. 36,7 százalékuk közép távú archiválást tervez, vagyis egy meghatározott időtartamig akkor is megőrzi és szolgáltatja a tartalmat, ha közben változnak a technikai körülmények, és szükség esetén megoldják a migrálást vagy az emulálást. Meglepő módon egészen magas, 43,3 százalék volt azoknak a repozitóriumoknak az aránya, amelyeknél a hosszú távú megőrzést jelölték be a válaszlehetőségek közül, vagyis „az idők végezetéig” szeretnék megoldani a digitális tartalom használhatóságát. Valójában ennek megvalósítása sok helyen még csak terv vagy vágyalom, legalábbis ez derült ki a szöveges hozzáfűzésekből és az utolsó kérdésre adott válaszokból, mert ezek szerint az intézményeknek mindössze 16,7 százalékánál megoldott a hosszú távú megőrzés.

Az összkép mindenesetre biztató – legalábbis ami az egyetemi szférában működő amerikai könyvtárakat illeti –, mert egyre több helyen foglalkoznak most már tudatosan a megőrzés problémájával: szabályzatokat és útmutatókat készítenek, kezelik a copyright-vonatkozásokat, minőségbiztosítással igyekeznek megőrzésre érdemes színvonalú és alkalmas formátumú tartalmakat összegyűjteni, és kihasználják az IR rendszerekbe beépített, illetve a külső szolgáltatók által nyújtott lehetőségeket a hosszú távú megőrzés érdekében. Mivel a legtöbb intézményi repozitórium egyelőre csak fejlesztési fázisban van, nyitott kérdés még ezek fenntarthatósága.

Hivatkozások

- [1] CONWAY, Paul: Preservation in the Digital World. Washington, D.C.: Council on Library and Information Science, 1996. március
<http://www.clir.org/pubs/abstract/pub62.html>

- [2] JACKSON, Julian: Digital Longevity: the lifespan of digital files. York: Digital Preservation Coalition.
<http://www.dpconline.org/events/previous-events/306-digital-longevity>
- [3] SMITH, Abby: Foreword to Census of Institutional Repositories in the United States MIRACLE Project Research Findings. Washington, D.C.: Council on Library and Information Science, 2007. február
<http://www.clir.org/pubs/reports/pub140/contents.htm#fore>
- [4] LYNCH, Clifford A.: Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. Washington, D.C.: Association of Research Libraries, 2003. február
<http://www.arl.org/bm~doc/br226ir.pdf>
- [5] BAILEY Jr., Charles W.: Institutional Repository Bibliography.
<http://digital-scholarship.org/irbl>
- [6] WATERS, Donald – GARRETT, John: Preserving Digital Information: Report of the Task Force on Archiving of Digital Information. Washington D.C.: The Commission on Preservation and Access, 1996. 68.
<http://www.clir.org/pubs/abstract/pub63.html>
- [7] KENNEY, Anne – BUCKLEY, Ellie: Developing Digital Preservation Programs: the Cornell Survey of Institutional Readiness, 2003–2005. 2005. augusztus 15.
<http://worldcat.org/arcviewer/1/OCC/2007/08/08/0000070519/viewer/file1088.htm#article0>
- /LI, Yuan – BANACH, Meghan: Institutional repositories and digital preservation: Assessing current practices at research libraries. = D-Lib Magazine, 17. köt. 5/6. sz. 2011./**

(Drótos László)